Budgeting and Planning for Data Management

March 2018

Introduction

Data Management is a task generally acknowledged as necessary but not always budgeted for explicitly within a project. Frequently data management is done by each researcher in a project, to the best of his/her ability, independently of other researchers in the project. There are advantages and disadvantages to this approach. One of the main problems is to ensure data management adheres to the same principles and is of the same standard throughout the entire project so as to ensure the data are suitable for archiving and publishing. Even if this is achieved by multiple researchers within a project, the process is more likely to be efficient if carried out, supported or co-ordinated by a data manager.

As with any task within the project, data management requires an allocation of time, money and other resources. There is also the need to ensure the appropriate skills are available to the research process.

The Data Management Role

The principal investigator needs to decide whether or not to appoint a data manager for the project. However, regardless of this decision, the functions of data management must be explicitly assigned to someone within the team. Appointing a data manager may be desirable when a project has multiple research processes going on simultaneously over a period of time, or when more than one researcher will require the support from someone who is technically competent with respect to managing the data.

The time and resources needed to effectively fulfil the data management tasks depend on the specific project size and complexity. In consequence, instead of giving an overall recommendation on the level of funding, we will establish the areas that we consider important to budget for.

The Tasks of the Data Manager

The data manager supports and contributes to the research team's effort to gather, clean, make available, process, store and publish the team's research data and accompanying documentation. He or she has the responsibility to implement any agreements established by the research team with respect to the data and to ensure the team adheres to good data management practices. He or she is also in charge of establishing a system of data quality checks. In a separate document in this pack we describe the Terms of Reference for a Data Manager but have mentioned some of the tasks here as a reminder of the elements of Data Management that need to be resourced:

- Implementation of the data management plan;
- Set up and maintain a data and document storage facility (DDS);
- Perform quality assurance checks on the data;
- Provide support to the research team on data management;
- Set up data entry systems;
- Provide input into training;
- Provide input into the design of data collection tools;
- Prepare data and documentation for archiving;

- Collate the metadata;
- Archive the data.

Time and Budget for other Team Members

Even if you have a full-time data manager, you will find that others in the team will also need to allocate some time to data management. In particular, the following areas should be included in your budgeting:

- Time for the principal investigator to oversee and co-ordinate the team data management efforts;
- Time for researchers to deal with data, data quality assurance, and data queries;
- Financial allocation for support staff or equipment for data entry;
- Financial allocation to establish a reliable system of backups. This might be to buy disk space on cloud servers, or it may be to buy and manage a more local backup system such as a local network drive. A combination of the two is not unreasonable.

Example list of Data Management Tasks

Between September 2010 and December 2012, a Data Management Consultant logged 120 days of data management work for CCAFS. This included the following tasks:

- Formatting the questionnaire for the Household Baseline survey for ease of completion and data entry;
- Allocating variable names throughout the questionnaire;
- Providing support for users in CSPro including creating demonstration videos;
- Revising the CSPro data entry system to match major changes in the questionnaire after the pilot phase;
- Writing the manual for the data entry system;
- Writing a Data Checking Guide for use with the CSPro data detailing how to produce frequency tables within CSPro;
- Writing a guide for further data checks within SPSS and producing the corresponding SPSS syntax; the guide included instructions on how to transfer data from CSPro to SPSS;
- Drafting the Analysis Plan document;
- Producing the SPSS syntax to accompany the analysis plan;
- Documenting the procedure for running the analysis plan syntax and extracting the results;
- Running quality control checks on the data;
- Merging cleaned data files and recoding and consolidation of crop and livestock codes;
- Creating a Dataverse for the CCAFS Baseline Study;
- Uploading files to the Dataverse;
- Working on the mitigation questionnaires including setting up a spreadsheet for data entry and carrying out the data entry and checking;
- Checking site analysis reports including re-running analyses where necessary;
- Creating training videos on completing the household questionnaire and using the data entry system;

- Contributing to the CCAFS Data Management Strategy;
- Producing an assessment report on the quality of the data from the Household Baseline Study;
- Producing a version of the data entry system with screen labels in Spanish for use in Nicaragua and other Central American countries;
- Producing guidelines for data management;
- Etc.

The CCAFS Household Baseline Study was a "large" study – the questionnaire used covered 20 pages which resulted in a total of 970 variables. The survey was run in 15 core sites across 12 countries with 140 households from each site. Thus, the resulting data file had 2100 records. For a study of this size, employing a full-time data manager is not unreasonable.

Summary

The debate about whether data management should be done by the researchers or by a dedicated data manager will no doubt continue. Some researchers have the capacity, the time and the inclination to do the data management themselves. However, in our experience, they are the exception. In most cases, support from a data manager who has been given explicit responsibility and authority to deal with data issues is essential to achieving the levels of quality that are expected from an international research effort.

Associate Videos

Videos accompanying the original release of the CCAFS Data Management Support Pack in 2013 are available as a playlist on the Statistical Services Centre YouTube Channel at https://www.youtube.com/channel/UCs7EU95YMjhvNozJKCD92xQ/playlists. These videos have not been updated since the original release but are mostly still relevant.

In particular the playlist includes a video on Planning and Budgeting for Data Management available from the following link:
https://www.youtube.com/watch?v=O0vpXLJPB5o&list=PLK5PktXR1tmNRaUPsFiYlyhg2lui0xgpj&index=4